



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### I might be fundamentally mistaken

**Citation for published version:**

Ridge, M 2015, 'I might be fundamentally mistaken', *Journal of Ethics and Social Philosophy*, vol. 9, no. 3, pp. 1-21. <https://doi.org/10.26556/jesp.v9i3.92>

**Digital Object Identifier (DOI):**

[10.26556/jesp.v9i3.92](https://doi.org/10.26556/jesp.v9i3.92)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Journal of Ethics and Social Philosophy

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## I Might Be Fundamentally Mistaken<sup>1</sup>

Quasi-realism aspires to preserve the intelligibility of the realist-sounding moral judgments of ordinary people. These judgments include ones of the form, “I believe that p, but I might be mistaken,” where ‘p’ is moral. The orthodox quasi-realist strategy (developed by Simon Blackburn) is to understand these in terms of the agent’s worrying that some improving change would lead them to abandon their belief.<sup>2</sup> However, it is unclear whether this strategy generalizes to cases in which the agent takes their error to be *fundamental* in a sense articulated by Andy Egan.<sup>3</sup> Egan suggests that Blackburn’s approach is the only game in town for the quasi-realist, and that its inability to handle worries about fundamental moral error therefore refutes quasi-realism tout court, and not just Blackburn’s version.

Egan’s challenge has generated considerable discussion, including an interesting reply by Blackburn, and further discussion by such influential theorists as Derek Parfit and Thomas Scanlon, both of whom endorse versions of Egan’s objection as especially telling against expressivism.<sup>4</sup> However, in my view we have not yet gotten to the heart of the matter. As Sebastian Köhler argues, the challenge can be reinstated in the face of Blackburn’s reply.<sup>5</sup> What is still needed is a fully general quasi-realist friendly theory of the nature of first-person judgments of fallibility, such that these judgments are demonstrably consistent with judging that the belief is stable – where Egan defines

---

<sup>1</sup> Many thanks to Alfred Archer, Matthew Chrisman, Andy Egan, Sebastian Köhler, Elinor Mason, Mark Schroeder and the anonymous referees who gave me useful feedback on a previous draft of this paper.

<sup>2</sup> See especially Simon Blackburn, *Ruling Passions* (Oxford: Clarendon Press, 1998).

<sup>3</sup> Andy Egan, “Quasi-Realism and Fundamental Moral Error,” *Australasian Journal of Philosophy*, 85 (2007): 205-219.

<sup>4</sup> Simon Blackburn, “Truth and *A Priori* Possibility: Egan’s Charge Against Quasi-Realism,” *Australasian Journal of Philosophy* 87 (2009): 201-213, Derek Parfit, *On What Matters* (Oxford University Press: Oxford, 2011), volume II, pp. 395-6, and Thomas Scanlon, *Being Realistic About Reasons* (Oxford: Oxford University Press, 2014), p. 60.

<sup>5</sup> Sebastian Köhler, “What is the Problem of Fundamental Error?” *Australasian Journal of Philosophy*, 2014.

‘stable’ in terms of the belief’s being such that no improving change would lead the agent to abandon the belief. In this article, I develop and defend such a theory, and argue that Egan’s challenge equivocates at a key point between a ‘could’ and a ‘would’.

### **I. Fundamental Moral Error: The Initial Challenge.**

Egan’s challenge begins with the idea that remarks of the form, “I believe that p, but I might be mistaken about that” are intelligible. This might not seem obvious, since there is something at least odd about voicing your belief that p and indicating doubts about p in the same breath. It can seem as if the speaker is taking back with one hand what he offers with the other. Egan has a useful discussion of this in a footnote (fn. 5), in which he distinguishes statements of the form “p, but I might be mistaken” from statements of the form, “I believe that p, but I might be mistaken about that.” While the former can seem odd, the latter seem perfectly acceptable and commonplace. He gives the nice example of a passenger checking to see if there are any oncoming cars telling the driver, “I don’t think there are any cars coming, but I might be mistaken.” Such examples are easily generated. The uncertainty signalled by simply saying one believes that p, rather than flat-out asserting that p, seems to make such remarks entirely unproblematic. The question, therefore, is not whether there is any data to accommodate here, but whether the realist has any advantage over the quasi-realist in accommodating it.<sup>6</sup> To see why this might be an issue we should first take a step back and get clear on how realism and quasi-realism differ.

---

<sup>6</sup> A further nuance is whether such remarks *still* are intelligible when the speaker adds that her belief is *stable* in Egan’s sense. This is hard to adjudicate by direct appeal to ordinary discourse since the notion of stability is an explicitly theoretical notion. However, the following comes close to capturing the relevant thought: “I believe that p, and I have excellent evidence for p which I doubt will be overturned by further investigation; still, I suppose I *might* be mistaken about p” Remarks of that form sound perfectly OK to my ear. In general, I think we should try to make room for people doubting even some of what they take to be their most well-grounded beliefs. The reasonableness (let alone mere intelligibility) of such doubts is, after all, something many philosophers have taken to be one of the deepest lessons of reflection on various forms of radical epistemological scepticism. David Hume famously and elegantly extolled the virtues of epistemic modesty based on sceptical reflection, e.g.

The moral realist characteristically *begins* with moral metaphysics, offering an account of moral states of affairs, and explains moral judgment in terms of *representing* these. The quasi-realist begins instead with moral psychology, focusing on their practical/desire-like functional role, and “earns the right” to the realist-sounding things ordinary people say. By appealing to a deflationist theory of truth, the quasi-realist accommodates moral truth. With truth, we have “taking true,” which is to say believing. By construing judgments of mind-independence as first order judgments about when something would be wrong, mind-independence is accommodated, and so on.<sup>7</sup>

It is therefore unsurprising that quasi-realists aspire to make sense of fallibility. Blackburn proposes that we understand what someone is up to when they worry about their own fallibility in terms of worrying that they might not live up to their own standards of judgment:

How can I make sense of my own fears of fallibility? Well, there are a number of things that I admire: for instance, information, sensitivity, maturity, imagination, coherence. I know that other people show defects in these respects, and that these defects can lead to bad opinions. But can I exempt myself from the same possibility? Of course not (that would be unpardonably smug). So I can think that perhaps some of my opinions are due to defects of information, sensitivity, maturity, imagination, and coherence.<sup>8</sup>

The thought is that worrying about whether you are mistaken is worrying about whether your judgments might be improved, where what counts as improvement by your lights is fixed by what you admire.

---

<sup>7</sup> For useful recent discussion, with a plausible diagnosis of why philosophers remain so tempted to attribute some form of problematic mind-dependence thesis to expressivists, see Sebastian Köhler, “Expressivism and Mind-Dependence: Distinct Existences,” *Journal of Moral Philosophy* (2013).

<sup>8</sup> Blackburn 1998, cited above, p. 318.

Egan argues that Blackburn's strategy breaks down when we stipulate that the agent's belief is taken by the agent to be *stable*. Unfortunately, Egan is not entirely consistent in how he characterizes stability. Here is Egan's canonical definition:

Call a belief *stable* just in case no change that the believer would endorse as an improvement would lead them to abandon it. Call belief *unstable* just in case it's not stable.<sup>9</sup>

Later, Egan operationally defines stability not in terms of whether a *single* change the believer endorses as an improvement would lead them to abandon it, but in terms of whether a *course* of "improving changes" would:

For my moral belief that P to be *stable* is for it to be such that it would survive any improving change (or course of improving changes).<sup>10</sup>

We need to clarify just what counts as a "course of improving changes." Unfortunately, 'course of improving changes' is multiply ambiguous and not defined by Egan. Each of the following has some plausibility:

- "ENDGAME STABILITY": No series of changes, each of which I would *at the end of the series*, given my epistemic standards *then*, regard as improvements would lead me to abandon the relevant belief.
- "FROM NOW STABILITY": No series of changes, each of which I would *now* regard as improvements, given my *current* epistemic standards, would lead me to abandon the relevant belief.
- "EACH STAGE STABILITY": No series of changes, each of which I would *at the time of the change* regard as improvements, given my epistemic standards *at that time*, would lead me to abandon the relevant belief.

---

<sup>9</sup> Egan, previously cited, 2007, p. 212.

<sup>10</sup> Egan, previously cited, 2007, p. 214.

EACH STAGE STABILITY is in my view the right reading. Crucially, if a moral belief is EACH STAGE STABLE, there will be no rational way for the believer to abandon it. This is important for Egan's challenge, but doesn't follow from FROM NOW STABILITY or ENDGAME STABILITY.

Recall that Blackburn's account holds that when one worries that one's belief might be mistaken, what one is worried about is that some series of improving changes might lead you to abandon it. By hypothesis, if a moral belief is (each-stage) *stable* then no series of improving changes would lead you to abandon it. So judging a belief is stable while worrying that it might be mistaken looks incoherent. Even worse, quasi-realists *can* make sense of the possibility that someone *else* might have a belief which is stable yet mistaken – crucially they need not endorse the other person's starting point. The quasi-realist therefore seems implausibly committed to others being open to a kind of error to which he has an a priori guarantee of immunity. Egan suggests that this attitude would make the expressivist come out as “unpardonably smug” (Blackburn's phrase) after all. This is not only implausible, it is of dubious coherence once we recognize that anyone else who accepts quasi-realism will then *correctly* infer that they too have the same special first-person a priori guarantee.

Blackburn's reply points out that Egan equivocates on 'improving change', as between “change which *is* an improvement,” and “change which the believer *would regard* as an improvement.” Here is Blackburn:

It is not quite right that for a belief that *p* to be stable is ‘for it to be such that it would survive any improving change (or course of improving changes)’. This gives a criterion of stability in terms of whatever *is* an improving change.

Whereas, as we have seen, officially stability is a matter of surviving anything that *the subject would regard as* an improving change, either antecedently, or

*post hoc*. Without this conflation, the result that a moral belief is mistaken only if it is not stable does not follow.<sup>11</sup>

Blackburn's point is well taken; Egan *does* slip between these two, and this undermines his challenge as stated. One might *argue* that, given expressivism, there is no distinction in the first person between what *is* an improving change and *what one regards as one*. Simply to *assume* there is no such distinction looks to beg the question, though.

Blackburn's point is sound as far as it goes, but it does not fully meet what is insightful in Egan's challenge.<sup>12</sup> What we really need is a positive account of what it *is* to judge that one's moral judgment might be mistaken, such that there no tension in simultaneously allowing that one might be mistaken about p while in the same breath asserting that the judgment that p is stable. Quasi-realists allow at the outset that they must *earn the right* to such realist-sounding remarks, so complaints about their critics begging the question are misplaced. Egan should have simply pointed out that Blackburn does not really tell us precisely what it is to judge that one might be fundamentally mistaken, at least not in a systematic and precise way, and with adequate generality. If you comb Blackburn's text to determine precisely what state of mind is supposed to be expressed by 'I might be mistaken' you will find no clear and fully worked out answer. Blackburn gestures at the idea that one is worried about traits one admires leading one to change one's view, but there are many ways to spell this out, and on many of them a version of Egan's challenge remains compelling, or so I shall argue below. Moreover, any such account must cohere with a plausible account of epistemic modals, something Blackburn does not discuss.

## II. Thinking One Might Be Mistaken

---

<sup>11</sup> Blackburn, previously cited, 2009, p. 205.

<sup>12</sup> See also Sebastian Köhler, "What is the Problem with Fundamental Error?" cited above. Köhler does not offer a solution to the problem, but argues in more detail than I can here that Blackburn's solution fails – for reasons similar to the one's I discuss in the text.

Independently of Egan's challenge, quasi-realists need an account of epistemic modals like 'might'. The sentence "It might not be the case that charity is morally good" is, after all, *both* a moral sentence *and* a modal sentence. If moral sentences get their meaning in virtue of expressing distinctive states of mind, then we need some account of what states of mind are expressed by such mixed moral/modal sentences. It will therefore not do to assume any old theory of epistemic modals will work.

Most discussion of the meanings of epistemic modals in the literature are couched at the level of semantics as opposed to meta-semantic theories, but it is not at all clear that expressivism is best construed as a semantic theory, rather than as a meta-semantic one. It is therefore useful at the outset to at least distinguish semantic theories from meta-semantic theories. Semantic theories provide a model of competence with natural languages, providing schemes for interpreting the messy sentences of natural languages in a more logically well-behaved meta-language. Such theories are characteristically recursive, providing a finite scheme for interpreting indefinitely many sentences. The most familiar versions of semantic theory in this sense are broadly truth-conditional, assigning extensions to the primitive predicates of the language, referents to singular terms, functions from context of utterance to contents (propositions), and so on.

Expressivism is sometimes taken to be an unorthodox semantic theory, on which semantic contents are not truth-conditions or propositions, but states of mind.<sup>13</sup> In my view, expressivism is better understood as a theory in meta-semantics, or what is

---

<sup>13</sup> For an explicit construal of expressivism as a view about semantic content in this sense, see Mark Schroeder, *Noncognitivism in Ethics*, 2012 (London: Routledge). For a dissenting view, see Alex Silk, "Truth-Conditions and the Meanings of Ethical Terms," forthcoming in *Oxford Studies in Metaethics*, volume 8. See also Michael Ridge, *Impassioned Belief* (Oxford: OUP, 2014), chapters 4 and 7 and Matthew Chrisman, "Attitudinal Expressivism and Logical Pragmatism," in Hubbs and Lind (eds.), *Pragmatism, Law and Language* (Routledge: 2014).



sometimes called the “foundational theory of meaning.”<sup>14</sup> A meta-semantic theory says in virtue of what a words and sentences have their meanings. My view is that quasi-realism construed as a meta-semantic theory is consistent with an orthodox truth-conditional approach to first-order semantics. Quasi-realism aims to vindicate not only talk of truth, but of the other semantic idioms (e.g., talk of the *extensions* of predicates) used in semantic theories. Obviously it is hotly debated whether quasi-realists *can* do all of this. Assuming that they can, though, quasi-realism (qua meta-semantic theory) is perfectly compatible with orthodox truth-conditional semantics.

Because I find meta-semantic versions of expressivism more promising, I here draw on my own account as developed in *Impassioned Belief* (Ridge 2014, cited above) in which a meta-semantic version of the theory is developed in more detail. My canonical formulation of the reply to Egan will therefore be couched in meta-semantic terms. It is important to note that this is not really essential to the reply, though. Insofar as the reply works at all, it should work for analogous semantic versions of quasi-realism which identify semantic contents as states of mind. So long as semantic versions of quasi-realism can make sense of the context-sensitivity of modals (which they had better be able to do anyway), this should not be a problem. Moreover, context-sensitivity in general should not be a problem for semantic forms of quasi-realism. Where traditional semantics understand the meanings of context-sensitive sentences as functions from contexts of utterance to propositions, a semantic form of quasi-realism can understand the meanings of such sentences as functions from contexts of utterance to states of mind. So long as the functions on offer are compatible with a broadly compositional theory of meaning, there should be no further objection to such semantic forms of quasi-realism on this front. Furthermore, it should be

---

<sup>14</sup> See especially Jeff Speaks, “Theories of Meaning,” *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/meaning/>.

obvious how to transpose the meta-semantic version of quasi-realism sketched here into a first-order semantic version of the view. In fact, such a theory is simpler, not needing to explain the relationship between the expression of states of mind and the expression of corresponding propositions. In effect, the theory is exactly the same up to the point at which the meta-semantic theorizing comes into play but stops at that point and instead simply identifies semantic contents of sentences in contexts of utterance as the states of minds expressed in those contexts. Bearing in mind the modularity of the solution developed here, I now return to the main line of argument and proceed in meta-semantic terms.

Plausibly, epistemic modals are context sensitive, and function to communicate that some proposition is *consistent* with contextually specified evidence (as with ‘might’), *necessitated* by that evidence (as with ‘must’) or *probable* in light of that evidence (as with epistemic ‘ought’). I favour a slightly non-standard account of how epistemic modals play these distinctive roles.<sup>15</sup> On this account, epistemic modals advert to what any acceptable epistemic standard would be like in some respect, given some contextually specified body of evidence. To a first approximation, ‘must’ adverts to what beliefs any acceptable epistemic standard would *require*, given the evidence, ‘might’ adverts to what beliefs any such standard would *permit*, and ‘ought’ adverts to what any such standard would *recommend*.

This account has several virtues. It lends itself to a broader account which provides sameness of meaning for modals across epistemic and deontic contexts. I develop this broader account elsewhere, but the basic idea is that deontic standards

---

<sup>15</sup> The key figure here is of course Angelina Kratzer, whose own views were to some extent inspired by the work of David Lewis. Kratzer’s work has by now been so influential that it would be silly to try to list those inspired by her work in a mere footnote. For a useful and up to date collection of relevant work, see Angelika Kratzer, *Modals and Conditionals* (Oxford: Oxford University Press, 2012). It should be obvious to those who know her work that my own account owes much to hers, even though it differs in important details too. Unfortunately, a full discussion of these differences would go beyond the scope of the present essay. See Ridge 2014 chapter 1 for further relevant discussion.

(moral standards, legal standards, standards of etiquette, etc.) can also either require or recommend, and uses of ‘ought’ and ‘must’ track this distinction there too. The account also explains why ‘must’ is logically stronger than ‘ought’ in both moral and epistemic contexts – roughly, requiring entails recommending but not vice-versa. It also explains why ‘must’ as used in deontic contexts is especially well-suited to issuing commands or requirements, while ‘ought’ is more naturally well-suited to offering advice. Much more could be said about this account, but here I am simply going to assume that it is roughly correct. My conclusion here is a modest one – Egan’s challenge can be met, *given* a semantics for modals roughly like this. It is worth remembering that Egan claims the force of his challenge does *not* depend on which of the reasonable semantic views about modals one favours:

The problem is *not* about giving a general semantics for ‘might’... (I do have views about these issues...but I don’t think they are particularly relevant to the issue at hand).<sup>16</sup>

Given this semantics, how does expressivism enter the frame? As a meta-semantic theory, expressivism offers a distinctive explanation of what it is in virtue of which words and sentences have their meanings. The basic idea is to explain meaning qua semantic content in terms of state of mind expressed, and then offer a distinctively non-representational theory of the states of mind for the target discourse. We can and should continue to speak happily of moral assertion, moral belief, etc. We simply need to allow that moral beliefs are different in their nature, and not just their content, from ordinary descriptive beliefs. I here single out moral and epistemic contexts, though.<sup>17</sup>

My explanation of why Blackburn’s reply does not really get to the heart of Egan’s challenge relies on the idea that, given the otherwise most plausible expressivist

---

<sup>16</sup> Egan, previously cited, p. 209, fn. 5.

<sup>17</sup> The extension of this approach to other uses of modals raises delicate issues I do not have the space to explore here. For discussion, see Ridge 2014, chapter 1.

treatment, the mental states which would constitute judging that one might be mistaken about *p* even though one's belief that *p* is stable involves some sort of incoherence. In my view, the best way to demonstrate why this worry is legitimate locates the incoherence in a straightforwardly inconsistent set of representational beliefs. It is not entirely clear how best to articulate this worry if normative judgments just *are* pro-attitudes, though, as Blackburn holds. If the contradiction arising is between the contents of representational beliefs, then we will need some way of associating some representational belief with the relevant normative judgment – that is, the judgment that “no improving change would...”

To be clear, I don't think that this allows Blackburn to sidestep the problem; it simply makes it harder to state. For example, pro-attitudes can commit one to various beliefs though devices like presupposition and these beliefs might contradict the stability belief. Perhaps my hope that the president is re-elected presupposes there is a president. In fact, though, it is hard to see how to develop Blackburn's suggestive remarks about normative uncertainty *without* effectively developing his account into a hybrid theory. For on his account the agent must not merely approve of some standards of judgment (coherence, sensitivity, etc.), he must also *believe* that his judgment might not live up to those very standards. The most natural way to turn Blackburn's theory into something more systematic therefore pushes strongly in the direction of a hybrid theory like Ecumenical Expressivism. In any event, it is easier to demonstrate the best version of Egan's worry in the framework of a hybrid or “Ecumenical” form of Expressivism.<sup>18</sup> This is also dialectically fair enough, as Egan's challenge is for *all* forms of quasi-realist expressivism, and not Blackburn's specific approach. Here is Egan:

---

<sup>18</sup> See e.g., Michael Ridge, “Finessing Frege,” *Ethics* 116 (2006): 302-336. See also Mark Schroeder, “Tempered Expressivism,” *Oxford Studies in Metaethics*, volume 8, ed. Russ Shafer-Landau (forthcoming), and Teemu Toppinen, “Believing in Expressivism,” also in v. 8 of *Oxford Studies in Metaethics* (forthcoming).

This leaves us with two options: abandon quasi-realism, or provide a different quasi-realist account of moral error... If it turns out that we can't provide such an account, we'll be forced to give up quasi-realism. I'm sceptical about the prospects for a substantially different quasi-realist account of moral error. Some story about stability under improving changes really does seem to be the best (perhaps the only) account of moral error that's available to quasi-realists.<sup>19</sup>

In effect, I am here taking up Egan's challenge to provide an alternative quasi-realist account of moral error – one developed within the framework of Ecumenical Expressivism.

Normative judgment is famously Janus-faced, having both belief-like and desire-like features. Ecumenical Expressivism tries to accommodate these features by understanding normative judgments as hybrid states. According to Ecumenical Expressivism, the relevant judgments are constituted by (roughly) a belief/desire pair, where 'belief' can be understood as denoting a state of mind with a robustly world-directed direction of fit.<sup>20</sup> Crucially, though, there is no *specific* representational belief which must constitute any given moral or epistemic judgment. Moral and epistemic judgments are on this view massively multiply realizable. What matters is not that the belief component of the judgment has any specific content. Rather, what matters is that the content of the belief-like component is *related* in the right way to the content of speaker's relevant pro-attitude.<sup>21</sup> This is why the view is meant to accommodate Moorean "Open Question Argument" style intuitions, and why the view contrasts with

---

<sup>19</sup> Egan, previously cited, pp. 216-217.

<sup>20</sup> For further discussion, see Michael Ridge, "Ecumenical Expressivism: Finessing Frege," *Ethics* 116 (2006): 302-336 and Michael Ridge, *Impassioned Belief* (Oxford: Oxford University Press, 2014).

<sup>21</sup> One might therefore with equal justice call the view a form of "relational expressivism." See Mark Schroeder, "Tempered Expressivism," cited above.

cognitivist hybrid theories which do privilege a representational content for each token normative judgment.

On the version of Ecumenical Expressivism I now prefer, such judgments are partly constituted by what I call the agent's *normative perspective*, which is a complex non-representational state of mind. In broad terms, they are diachronically stable high-level *intentions* not to endorse certain kinds of standards, and to act and deliberate only in ways permitted by those standards not ruled out. For present purposes, though, the crucial features of normative perspectives are that (1) they are non-representational, and (2) they *rule out* standards (both moral and epistemic) of certain kinds. Here "ruling out" can be understood in terms of the agent setting themselves against standards of a certain sort – firmly intending not to endorse them. This is less immediately relevant here, but they also rationally commit the agent to acting, deliberating and reasoning only in ways that would not be forbidden by all such standards. Normative perspectives are, then, complex high level intentions of a certain sort. In fact, the details of the account of normative perspective are not essential to the reply to Egan developed here. All that is essential is that they have a desire-like direction of fit and function to rule out certain standards of practical reason. Any version of hybrid expressivism rich enough to have these structural features should be able to deploy the solution on offer here.

On this sort of account, a given agent's moral or epistemic judgment will then be partly constituted by a normative perspective and partly constituted by a robustly representational belief whose content is related in the right way to that perspective. Consider my judgment that morally I must give to charity. On the proposed account, this judgment will be constituted by a normative perspective and the belief that any admissible moral standard (any standard not ruled out by *that* perspective) would

require my giving to charity. I here use ‘standard’ in a maximally inclusive way to include all of an agent’s moral principles at a given time.

I favour a similar account of epistemic judgments, including judgments made using epistemic modals. This brings us to my proposed account of first-person judgments of (moral) fallibility. I can now explain why the framework of Ecumenical Expressivism is useful here. The real force of Egan’s challenge is the worry that a judgment of first-person fallibility about *p* somehow contradicts or stands in tension with a believer’s at the same time being *certain* that her judgment that *p* is stable. The worry is that there may well be some descriptive belief associated with a judgment of fallibility – which for the expressivist is a normative judgment – and that this ordinary descriptive belief contradicts or stands in tension with the relevant stability belief. Ecumenical Expressivism makes it easier to state the challenge in a clean way because it provides a *very* clear sense in which a token descriptive belief will be associated with a token normative judgment. On that view, any token normative judgment will be partly constituted by some such descriptive belief. Call any such token belief the “descriptive realizer” of the fallibility judgment. The issue is whether the descriptive realizer contradicts the stability belief. For a non-ecumenical expressivist, like Blackburn, the issue would be whether some belief presupposed by the normative judgment contradicts the stability.

Is there any reason to think that the relevant beliefs are contradictory? Yes. Consider how the orthodox approach from Blackburn would look if transposed into the framework of Ecumenical Expressivism. The simplest transposition would understand judgments of fallibility as constituted by (a) the admiration of certain traits, and (b) some belief corresponding to improvements as defined by those traits. The problem with this approach is that it does not give any guidance as to what the agent

takes as an improvement, all things considered, when the varied traits she admires pull in opposite directions. Since we need some account that is general enough to make sense of generic judgments about what improving changes quite generally would yield, we need to refine Blackburn's account.

A tempting modification is to understand such judgments in terms of the agent's standards for judgment. Because the relevant standards can include whatever priority orderings the believer endorses, we are not left in the dark about how to understand the agent having a view about what would be better in the relevant sense when different desiderata conflict. This modification can capture what is plausible in Blackburn's approach, since one might endorse a standard because it embodies traits one admires. In the following section I explain how a version of Egan's challenge can still arise in this context. I then explain how this remaining challenge can be met, and how this in turn provides a deeper diagnosis of how Egan's argument goes wrong.

### **III. Meeting the Challenge: 'could' versus 'would'**

I begin this section with an important caveat. The account I develop in this section makes free use of the idea of assigning credences to normative propositions.

However, it is not at all obvious how an expressivist should understand talk of such credences, or indeed whether such an account is even possible. There is an extensive debate on this topic initiated by Michael Smith's classic paper, "Evaluation, Uncertainty and Motivation,"<sup>22</sup> which argues that expressivists do not have the resources to distinguish three features of normative judgment we independently have good reason to keep distinct – importance, robustness and certitude. Unfortunately,

---

<sup>22</sup> Published in *Ethical Theory and Moral Practice*, 2002: 305-320. For some attempts to meet the challenge, see my "Certitude, Importance and Robustness for Non-Cognitivists," <http://www.brown.edu/Departments/Philosophy/bears/0301ridg.html>, James Lenman's post, "Non-Cognitivism and the Dimensions of Evaluative Judgment," <http://www.brown.edu/Departments/Philosophy/bears/0301lenm.html>, and Michael Ridge, "Ecumenical Expressivism: The Best of Both Worlds," in *Oxford Studies in Metaethics* (ed. Shafer-Landau), 2007, v. 7. For replies to those solutions, see Krister Bykvist and Jonas Olson, "Expressivism and Moral Certitude," *Philosophical Quarterly*, 2008: 202-215.



an attempt to develop an adequate expressivist theory of credences in normative propositions raises very complex issues which go well beyond the scope of this article. Indeed, even to engage properly with the existing complex literature on this difficult topic, much less lay out and defend a positive proposal, would easily require an entire article in and of itself. Here I must simply assume that some such theory is in the offing. I recognize that this is no trivial commitment. My main aim in this paper is, in effect, to reduce these two problems into a single problem.

If I can reduce these two problems to one problem, though, this strikes me as intellectual progress in at least two respects. First, if I am right about the state of the dialectic prior to the present article, it was in no way obvious that a defensible expressivist theory of talk of normative credences would be *sufficient* for the expressivist to make sense of the possibility of fundamental normative error. Insofar as we have reason to be optimistic that *some* expressivist theory of credences in normative propositions will turn out to be defensible, the argument of the present paper, if sound, provides reason to be just as optimistic about an account of fundamental normative error. Establishing that confidence about the expressivist theory of normative credences is enough for similar confidence about the solution to Egan's problem is therefore one important lesson of this paper.<sup>23</sup>

---

<sup>23</sup> As it happens, I am cautiously optimistic that Ecumenical Expressivism, as developed in Ridge 2014, has the resources to provide an adequate account of normative credences – or, at least, a more promising account than has so far been defended in the literature. The hybrid structure of the theory offers resources for distinctions not available to more traditional forms of expressivism. More importantly, the main problems (raised by Bykvist and Olson) with my previous account arguably arose from features peculiar to my earlier “ideal advisor” version of the account. In particular, the move to “normative perspectives” offers some new and interesting resources for thinking about credences in normative propositions. Here it matters that normative perspectives are constituted by two very different kinds of intentions, one of which is negative (not to adopt certain standards) and one of which is positive (to act and deliberate only in ways allowed by all standards not ruled out by the perspective's negative component). This suggests new strategies for developing a theory of credences which does not gloss them in terms of *relative* motivational strength (as with my previous account), such that an increase in one's non-normative motivations (Bykvist and Olson give the example of falling in love) absurdly entails a change in an agent's confidence in all of her fundamental normative judgments. Instead, the newer version of Ecumenical Expressivism has resources for a non-relative theory of fundamental normative credences which can define an interval from 0 to 1 independently of

A second lesson is that we do in fact *need* an expressivist theory of normative credences. If I am right that the best strategy for meeting Egan’s challenge makes free use of the idea of assigning such credences, then one response to Smith’s challenge should to that extent off the table. Here I have in mind those who suggest that the very idea of assigning credences to fundamental normative propositions makes no sense – that is, an error theory about such credences. Such a theory does, after all, have at least some *ex ante* attractions. For one thing, as James Dreier has argued, it is not clear what the relevant counterfactual betting behaviour which should be associated with such credal assignments could *be*.<sup>24</sup> If the argument of this paper is correct, then the expressivist *had better* be able to avoid such an error-theoretic conclusion, on pain of not being able to dissolve Egan’s problem. Insofar as it was not previously obvious that such a theory was independently necessary to defend expressivism from other objections (previous replies to Egan do not draw on the theory of credences, after all), this is a second way in which the account developed here should be instructive even if it is not the final word on these complex matters.

Given the framework of the preceding section, judgments of the form, “It might be the case that not-p” quantify over standards. They can therefore be paraphrased in terms of what any acceptable epistemic standard would be like in some relevant way. A relevant difference between my account and Blackburn’s is that my account is not couched in terms of what correctly following the relevant standards would *lead to*. Instead, it is couched in terms of what the relevant standards would *permit*. Since I

---

the agent’s other motives and do so in a principled way for each fundamental normative proposition the agent has a view about. It also seems to me that it can do so in a way that avoids some of the other problems Bykvist and Olson raised for non-relative conceptions in general. However, it would simply take me too far afield from the main themes of the present paper to elaborate on these unfortunately very elliptical remarks here. The theory of normative credences itself, and the details of my more recent work needed for the backdrop to that theory, not to mention the dialectic with Bykvist and Olson’s numerous objections quickly becomes extremely complex.

<sup>24</sup> See Dreier’s discussion on the “Pea Soup” ethics blog here:

<http://peasoup.typepad.com/peasoup/2009/01/expressivism-and-moral-certitude.html>.

might not be disposed to take advantage of any given permission, this might seem sufficient to deal swiftly with Egan's challenge. Compare: "I am legally permitted to scream, but correctly following the law would not lead me to scream." However, this reply is too swift. We can reinstate the challenge by redefining stability in terms of how one's beliefs are *permitted* to evolve.

I now need to refine the semantic theory sketched above, itself put forward as a first approximation. On that rough version of the theory, 'must-p' corresponds to 'any acceptable epistemic standard would, given some contextually specified body of evidence, require believing that p'. One problem with this is that it is not clear that acceptable epistemic standards ever *require* believing any given proposition. Whether to form a belief one way or the other on a given topic seems like a pragmatic question, and not one that would be answered by purely epistemic standards (even given some level of what epistemologists call "pragmatic encroachment"). A more plausible version of the theory would hold that 'must-p' corresponds instead to 'any acceptable epistemic standard would, given some contextually specified body of evidence, require assigning a much higher credence to p than to not-p *if one assigns them credences at all.*' This leaves it open whether one should form a view about p at all. We can then follow orthodox deontic logic, and understand 'might-not-p' as equivalent to 'not-must-p'. So 'might-not-p' will correspond to "it is not the case that any acceptable epistemic standard would, given some contextually specified body of evidence, require assigning a much higher credence to p than to not-p if one assigns them credences at all." Assuming not requiring is permitting, 'might-p' corresponds to "some acceptable epistemic standard would, given a contextually specified body of evidence, permit not

assigning a much higher credence to *p* than not-*p*, if one assigns them credences at all.”<sup>25</sup>

Note that I have defined ‘must’ in terms of assigning a *much higher* credence to *p* than not-*p*. This is essential to the plausibility of the proposal. One might have thought that ‘must’ should instead be glossed in terms of absolute certainty – assigning a credence of 1 to the proposition in question. Indeed, this would in some ways make for a more elegant theory. In fact, though, even a moment’s reflection on the arguments for radical epistemological scepticism, many of which have entered the public consciousness now – think e.g. of cultural influence of *The Matrix*, demonstrates how problematic this approach would be. Ordinary speakers who readily admit that they cannot be certain that they are not “in the Matrix” will say things like “Jones must be home by now.” Even putting such radical sources of doubt to one side, ordinary speakers regularly say propositions must be true which they clearly could not reasonably assert can be known with the same level of certainty as  $2+2=4$  or the validity of modus ponens. “He must be winning, he is up a pawn for nothing,” for example, is the sort of remark one might make about a chess game without being willing to assert that you are or should be *absolutely* certain that the person in question is winning. Given the ubiquity of such contexts, it would in my view be implausibly heavy-handed to insist that these speakers are either confused, do not really mean what they say or are

---

<sup>25</sup> I realize that the assumption that a standard’s permitting something is identical to its not requiring it is debatable. Here it is important to bear in mind that ‘standard’ is really a technical term for a comprehensive set of the relevant sorts of principles. This is important because a standard in this sense can be thought of as analogous to a total legal regime, and it is plausible that if no law in an entire legal regime *L* requires  $\Phi$ -ing in *C* then  $\Phi$ -ing in *C* is permitted by that regime. Whereas if one thought of standards in the present context as more analogous to individual laws within a given regime then the transition from ‘not required by this standard’ to ‘permitted’ would be very dubious indeed. Obviously a lot more would need to be said about how we should understand talk of standards in the framework of Ecumenical Expressivism to fully vindicate the transition made in the text, but this would take us too far afield here. For some further discussion, see Ridge 2014: 25-26 and 129. Thanks to an anonymous referee for drawing me out on this point.

only “speaking loosely.” Hence my account’s formulation in terms of assigning a *much higher* credence rather than in terms of assigning a credence of 1.<sup>26</sup>

So far, so semantic. At the level of meta-semantics, we explain how sentences deploying epistemic modals get their meaning in terms of their expressing hybrid states. Recall that the relevant judgments are understood as partly constituted by a normative perspective, and partly constituted by a belief whose content is fixed by the content of that perspective. So ‘it might be the case that not p’ gets its meaning in virtue of expressing (a) a normative perspective, and (b) the belief that some admissible epistemic standard (that is, some such standard compatible with the speaker’s normative perspective at the time of utterance), would, given some contextually specified body of evidence, permit not assigning a much higher credence to p than not-p, if one assigns them credences at all.

Given this framework, the question is whether this descriptive realizer contradicts belief in the relevant form of stability. Stability is now defined in terms of whether the relevant belief would be deemed *permissible* after any series of improving changes. In order to maximize the chance of a conflict with the descriptive realizer, we should define stability in terms of credence-assignment too:

---

<sup>26</sup> At the same time, ‘much higher’ is *also* context-sensitive, and how much higher the credence must be to count as ‘much higher’ will therefore also depend on the context. This is important because it may well be that in *some* contexts, ‘must p’ can only be properly asserted if the speaker’s favoured standards all require assigning a credence of 1 to p and a 0 to not-p (if one assigns it a credence at all) because only such a wide gap will count as ‘much higher’ in the contextually specified sense. I suspect that setting the bar so high may be very unusual, but perhaps it can happen. For example, this may be the standard in the philosophy seminar room in which various forms of radical scepticism are very much on the table, for example. In this respect, my account here is compatible with moves in the spirit of those that David Lewis famously makes in “Elusive Knowledge.” See David Lewis, “Elusive Knowledge,” 2006, *Australasian Journal of Philosophy*, 74: 549-567. The same point applies to ‘might p’. This is important because it helps explain why ‘might p’ can sometimes still be properly asserted even when all of the favoured standards require assigning what would count as a “much higher” credence to not-p on most of the more normal contextual specifications of the content of ‘much higher’. Still, so long as some contexts can set the standards for ‘much higher’ so that only the maximum possible gap counts as ‘much higher’, there will be contexts in which such remarks are properly asserted even though not-p is properly assigned a very high credence. Thanks to an anonymous referee for drawing me out on this nuance – which I take to be a general issue for all theories of epistemic modals which try to respect the fact that ordinary speakers do not typically require absolute certainty about p for asserting ‘must p’.

**STABILITY\***: Agent A's assignment at time t of a much higher credence to p than not-p is (by definition) stable if and only if that assignment is such that any epistemic standard which A would, given his normative perspective, correctly deem to be acceptable, would, given A's evidence, classify A's assignment of a much higher credence to p than not-p as permissible, *and* any series of improving changes (as defined by A's standards at the time of the improvement) is such that it would not overturn the classification of A's assignment of a higher credence to p than not-p as permissible.

Does agent's belief that p being **STABLE\*** somehow contradict the descriptive realizer? The relevant realizer has the following content:

**DESCRIPTIVE-REALIZER**: Some admissible epistemic standard (that is, some standard which the speaker would, given his normative perspective, correctly deem to be acceptable) would, given contextually specified evidence, permit not assigning a much higher credence to p than not p, given that one assigns them credences at all.

However, a careful reading reveals that a judgment of **DESCRIPTIVE-REALIZER** for a given agent about her belief that p does not *quite* contradict a judgment by that same agent at the same time of **STABILITY\*** about her belief that p. These are consistent because they both issue *permissions*. So long as the admissible standards mentioned by **DESCRIPTIVE REALIZER** not only permit not assigning a higher credence to p than not-p, but *also* permit assigning a higher credence to p than not-p, given the same body of evidence, **STABILITY\*** and **DESCRIPTIVE REALIZER** are logically consistent.

We need to invoke the following auxiliary hypothesis to derive a contradiction:

**EPISTEMIC PRESCRIPTIVITY:** No acceptable epistemic standard would, for any given body of evidence, both permit assigning a much higher credence to *p* than not-*p* and also permit not assigning a much higher credence to *p* than not-*p*, given that one assigns them credences at all.

I will not offer a detailed argument for **EPISTEMIC PRESCRIPTIVITY**, but simply highlight its plausibility. It is a corollary of the familiar idea that one must proportion one's beliefs to the evidence. If the evidence strongly favours *p* more than not-*p*, one is required to assign a much higher credence to *p* than not-*p* *if* one assigns any credence to them. If the evidence is equally balanced between the two, this does not mean one can willy-nilly assign a much higher credence to whichever one prefers.<sup>27</sup>

Given **EPISTEMIC PRESCRIPTIVITY**, a contradiction between **STABILITY\*** and **DESCRIPTIVE REALIZER** follows. **DESCRIPTIVE REALIZER** entails there is at least one admissible standard which permits not assigning a much higher credence to *p* than not-*p*, given that one assigns them credences at all. **EPISTEMIC PRESCRIPTIVITY** entails, by the believer's lights, anyway, that this standard does not also permit assigning a much higher credence to *p* than not-*p*. We have stipulated that the standard in question is deemed acceptable by the believer in question. **EPISTEMIC PRESCRIPTIVITY** entails that no acceptable standard which permits assigning a much higher credence to *p* than not-*p* also permits (relative to the same body of evidence) not assigning a much higher credence to *p* than not *p*, given that one assigns them a credence at all. However, **STABILITY\*** entails that *all* standards which the believer deems acceptable *do* permit assigning a much higher

---

<sup>27</sup> One worry about **EPISTEMIC PRESCRIPTIVITY** is that if some form of "pragmatic encroachment" is correct then acceptable epistemic standards might recommend different credences in a proposition depending on more than the evidence – in particular, depending on the practical stakes. If this is correct, though, it just requires a reformulation of the challenge in the text, for in this case the relevant epistemic standards will also need to advert to the stakes. So both **DESCRIPTIVE REALIZER** and **EPISTEMIC PRESCRIPTIVITY** would need to be reformulated so as to advert to both evidence *and* stakes. Once they are both reformulated along the same lines, though, the argument still goes through. Thanks to Matthew Chrisman for useful discussion.

credence to p than not-p. Thus the same standard both permits and does not permit assigning a much higher credence to p than not p; contradiction.

How might this challenge be met? Recall that to get a content epistemic modals require a contextually specified body of evidence (a “modal base”). A crucial question is how this modal base is construed in the context of someone worried some stable belief is mistaken. Clearly, such a believer is not worried that her *actual* evidence would rationally lead her to revise her view; its stability rules this out. Neither, given stability, is such a believer worried that the evidence she would in fact uncover, if she diligently sought more evidence, would rationally lead her to revise her view.

What we need is a modal base which ranges well beyond the evidence the believer has or considers likely to arise through further investigation. Indeed, given that she is worried about fundamental error in the face of robust stability, plausibly she is worried about epistemic scenarios which she has not entirely ruled out, but which she considers extremely unlikely. This suggests that a charitable interpretation would specify the modal base in terms of any possible body of evidence *compatible with everything the believer has not completely ruled out* – i.e., not assigned a credence of *zero*.

Thinking about the kinds of moral propositions with regards to which one might worry about fundamental moral error should help bring out why this is a plausible interpretation. Clearly the relevant beliefs will need to be fundamental moral commitments; otherwise I might just be worried about factual error. Consider my belief that pleasure is sometimes good as an end. I might well take this belief to be stable, yet allow that it might be mistaken. Perhaps I cannot even imagine a possible body of evidence which would rationally lead me to reject the proposition that pleasure is sometimes good as an end. However, I realize that this may reflect a failure of



imagination on my part, rather than deep insight. I therefore allow that there *may* be some very distant possible world in which I encounter evidence which, given my standards, would rationally permit me to abandon this belief. Intuitively, this seems to fit well with how someone might actually think about the possibility of error about such fundamental commitments. On the proposed account, my allowing for such epistemologically distant possible worlds will be constituted by (a) my normative perspective, and (b) the belief that relative to some body of evidence (where bodies of evidence are construed broadly enough to include deliverances of a priori reflection) B, such that not all admissible epistemic standards require assigning a credence of zero to B, come what may, and such that, given B, all admissible epistemic standards would permit rejecting the belief that pleasure is sometimes good as an end.

This provides the key to meeting the challenge. Crucially, stability is defined in terms of what *would* be the case if the relevant epistemic standards *were* followed in light of further information or reflection. Stability therefore gives us a counterfactual conditional. To assess counterfactual conditionals we examine only the relevantly *nearby* worlds. The belief which partly constitutes the relevant sort of first-person judgment of fallibility (the descriptive realizer), by contrast, ranges over a much wider set of worlds – *all* of those epistemic worlds (possible bodies of evidence) not ruled out with certainty, come what may, by the agent. My belief that p might be stable, in that I would not *in fact* come to the conclusion that it is permissible to abandon that belief, and this might be so even if my application of my epistemic standards were informed by arbitrarily more relevant evidence, greater imagination, more care, etc. This, though, does not entail that there are not also scenarios compatible with everything I have *not completely ruled out*, such that in those scenarios following my epistemic standards correctly would lead me to the conclusion that my belief is permissible to abandon.

Consider my toy example again. I might be *very* confident that if I were to ask a moral saint whose testimony I give enormous weight, he would not tell me pleasure is not sometimes good as an end. However, I might not be confident enough about this to assign it a credence of zero.

If this distinction seems obscure, consider a simple example. I am extremely confident that if I were to drop my pen then it would fall to the floor. However, I allow that it *might* be the case that I drop my pen and it doesn't fall. These judgments are consistent. Why? The first concerns the *nearby* worlds in which I drop my pen; the second concerns the wider set of all the worlds I have not ruled out *with certainty* in which I drop my pen. There are *distant* but epistemically possible worlds I do not reject with certainty in which a fairy keeps my pen afloat. Precisely because these worlds are distant this does not contradict my belief that the pen would fall if dropped.

My proposal is that the stability belief is analogous to the counterfactual conditional, "If I were to drop my pen, then it would fall to the floor," while the descriptive realizer for my "I might be mistaken" judgment is analogous to "it might be the case that I drop my pen and it doesn't fall to the floor." This is why the apparent force of Egan's challenge ultimately trades on a failure to distinguish a 'would' from a 'could'. It is only by paying careful attention to the semantics for epistemic modals that we can see both how a residual version of the challenge remains, given **EPISTEMIC PRESCRIPTIVITY**, yet ultimately rests on a failure to distinguish a 'would' from a 'could'. Once this distinction is carefully heeded, even the most powerful remaining version of the challenge dissolves.

At this point someone sympathetic to Egan's general strategy might reasonably complain that my reply, while effective against his actual argument, would have no force against a slightly revised version. Egan simply erred, one might suppose, in formulating

stability in counterfactual terms. He should instead have formulated it in terms of the *impossibility* of a series of improving changes leading the agent to revise her judgment. In that case, it seems, there will be no equivocation between a ‘would’ and a ‘could’ and the argument could be reinstated. Prima facie this is a very plausible strategy for salvaging Egan’s argument, but it does not quite work. It is instructive to see precisely why it fails.

The revised argument would replace STABILITY\* with HYPER-STABILITY:

**HYPER-STABILITY:** Agent A’s assignment at time t of a much higher credence to p than not-p is (by definition) stable if and only if that assignment is such that any epistemic standard which A would, given his normative perspective, correctly deem to be acceptable, would, given A’s evidence, classify A’s assignment of a much higher credence to p than not-p as permissible, and *no possible* series of improving changes (as defined by A’s standards at the time of the improvement) *could* overturn the classification of A’s assignment of a higher credence to p than not-p as permissible.

The crucial shift, of course, is that the second disjunct of STABILITY\* was a counterfactual, whereas the second disjunct of HYPER-STABILITY is an impossibility thesis. Now, it seems, the content of DESCRIPTIVE REALIZER will indeed contradict the content of HYPER-STABILITY. In that case, though, if the version expressivism developed above is correct, then anyone who simultaneously judges that her moral belief that p is both hyper-stable and that she might be mistaken about p will thereby have contradictory beliefs. Nor will it be contradictory to judge that someone *else’s* moral belief might be mistaken in spite of being hyper-stable, since of course the agent need not endorse the standards in virtue of which the other person’s judgment is

hyper-stable. Once again we have an a priori guarantee against a kind of error we can at the same time intelligibly attribute to others.

Indeed, the argument from hyper-stability is a significant improvement on Egan's original argument for precisely this reason, and it requires a different reply. The crux of my reply to this version of the argument is that while anyone whose substantive moral judgment is hyper-stable would thereby be "unpardonably smug," the fault here lies with the agent and not with the expressivist account of her judgment. That is, *simply* being an expressivist will not be sufficient for you to correctly judge that any of your actual moral judgments are immune from error in this privileged way. You must also deem the judgment in question to be hyper-stable. Given just how strong hyper-stability is, though, insofar as you correctly judge one of your substantive moral beliefs to be hyper-stable you thereby manifest an epistemic vice – indeed a vice we might reasonably characterize as smugness.

For **HYPER-STABILITY** to combine with **DESCRIPTIVE REALIZER** to entail a contradiction, the modalities must be the same. The modality invoked by **DESCRIPTIVE REALIZER** is clearly *epistemic*, though, as it is explicitly relativized to a contextually specified body of evidence. Here we are interested in what is possible given any body of evidence the agent has not entirely ruled out, of course. In that case, in order to generate a contradiction, **HYPER-STABILITY** must assert that no merely *possible* body of evidence could undermine the relevant judgment, where being possible is simply a matter of being consistent with whatever the agent has not entirely ruled out – that is, assigned a credence of *zero*.

The key premise of my reply to the revised argument is that we should not be this dogmatic about even our most confident substantive moral judgments. Note: the caveat 'substantive' is important here –obvious moral tautologies raise special issues I

discuss separately below. Clearly moral judgments which rely on empirical assumptions fail this test. My confidence that what Hitler did was morally wrong is of course dependent on my knowledge of what Hitler did, but I cannot reasonably assign a credence of 1 to that proposition. After all, I cannot even assign a credence of 1 to the proposition that I am not a brain in a vat, which would undermine my confidence that Hitler even existed. The more interesting judgments are therefore fundamental moral judgments – judgments which do not rely on contingent empirical assumptions.

Take, for example, my judgment that pleasure is sometimes good as an end. I am *very* confident about this, but can I reasonably assign it a credence of one – and thus assign its negation a credence of zero? In my view, I cannot. For I cannot reasonably assign a credence of zero to *nihilism* – the thesis that nothing is really good in any robustly normative sense of ‘good’. For one thing, even if I am very confident about my expressivism, I certainly cannot assign it a credence of one, nor can I assign any of its main rivals in the literature a credence of zero. After all, I know that several of my very smart peers have developed *prima facie* powerful arguments for the so-called “error theory” about the normative tout court. It would be arrogant (unpardonably smug?) for me to assign these argument so little weight that I give literally *no* credence to the error theory. Insofar as any such error theory is true, though, *nothing* is ever good as an end – so *ipso facto* pleasure is never good as an end. So I cannot reasonably assign a credence of zero to the proposition that pleasure is not ever good as an end. The same argument generalizes to any of my most firmly held basic moral or normative beliefs.

Nor is the point limited to epistemic modesty at the meta-normative level. Even as a matter of first-order normative theorizing, I cannot reasonably assign nihilism a credence of zero. Perhaps in seeking a Rawlsian “wide reflective equilibrium” I would

be forced to conclude that nothing is really good as an end. I deem this very unlikely, but it would again be arrogant to rule out the possibility entirely. The epistemic possibility of nihilism as a first-order view is therefore also sufficient to make the point. Nor, indeed, is the point limited to nihilism. Although I do confidently think pleasure is good as an end, there are rival non-nihilistic normative theories worth taking seriously which entail that this is false. Certain forms of deontology might hold that only the good will is really good as an end. Kant himself did not think this (he allowed that pleasure was sometimes good as an end, only not unconditionally so), but there is nothing in the Kantian framework which seems to *require* supposing that pleasure is ever really good as an end. Certain intelligible forms of Stoicism and ascetic normative perspectives might also deny this. In my view, the same point will apply to any fundamental normative judgement.

The point, then, is that we should recognize the fallibility of even our most cherished moral and other normative judgments. Obviously a lot more could be said about this, but this is not the place for a broader discussion of fallibilism in epistemology. It will be enough for my purposes if I can show that Egan's challenge can be met so long as the expressivist can appeal to the form of fallibilism I have sketched. It might be worth noting that it would be dialectically awkward for someone pressing Egan's form of argument to deny fallibilism. After all, the point of the Egan argument in a way is to make room for a robust and fully general kind of fallibilism.

What, though, about obvious moral tautologies, like "pleasure is either morally good or it is not." I do not want to deny that we cannot assign a credence of one to these, nor that we can assign a credence of zero to their negations. In this case, though, I think the right thing to say is that *everyone* should assign a credence of zero to obvious contradictions. So while I cannot intelligibly wonder whether I am mistaken about

some obvious moral contradiction, I cannot intelligibly think anyone *else* might be mistaken in their denial of an obvious moral contradiction either. Obvious contradictions are a special case in this way. Since there is no asymmetry between myself and others here, nor is there any reasonable worry about my being smug, the argument cannot be reintroduced by focusing on the special case of obvious moral tautologies and contradictions.

There is one last move a defender of Egan might make at this stage, though. Granted, insofar as I am not overly dogmatic, none of my *actual* moral judgments will be hyper-stable. Still, I can still think about what *would* be true if one of my judgments *were* to become hyper-stable. Perhaps expressivism entails that I must judge that *if* one of my substantive moral judgments *were* to become hyper-stable *then* I could not be mistaken about it, and this will allow Egan's challenge to be reintroduced. I hope it is obvious why this strategy is hopeless. Blackburn's original reply to Egan (discussed in section I above) is highly germane at this point. For when thinking about whether I might be mistaken if I were to become more confident about some moral question here and now I will still be deploying my *actual* epistemic standards – and not the much more dogmatic epistemic standards I am imagining my hypothetical self to have. I will therefore rightly view this hypothetical version of myself as lacking in epistemic virtue and therefore *inappropriately* confident in his moral judgment. Nothing in the expressivist theory commits me to endorsing any merely possible set of epistemic standards I might imagine myself to have. The point is precisely the same one that Blackburn and others have made time and time again in explaining why quasi-realist expressivism is not a form of subjectivism, and does not entail that the relevant class of judgments involves a problematic form of mind-dependence.

The overall form of my reply to the Egan-style argument then is as a dilemma. Either Egan's argument is couched in terms of stability or hyper-stability. If it is couched in terms of mere stability, then the argument equivocates between a 'would' and a 'could'. If, on the other hand, the argument is couched in terms of hyper-stability then no such equivocation follows. In that case, though, the argument fails for a different reason. For merely being an expressivist does not provide one with a problematic a priori guarantee of immunity from error. Rather, one must *both* be an expressivist and at the same time make substantive moral judgments which are hyper-stable. However, I have argued that the hyper-stability of substantive moral judgments *itself* manifests an epistemic vice – and, indeed, a vice one might reasonably gloss as being “unpardonably smug.” In that case, though, the conclusion of the argument would only be that being both an expressivist and independently of one's expressivism being unpardonably smug entails that you are unpardonably smug. That, though, is hardly an objection to expressivism!

Note, moreover, that the contrast between stability and hyper-stability is important here precisely because merely stable substantive moral judgments need not entail any sort of epistemic vice. This is why my main reply to Egan's actual argument is necessary to the argument as he actually develops it – on my reading, anyway. The broader point, then, is that the defender of Egan's form of argument must choose their poison. Either formulate the argument in terms of stability and then fall prey to the conflation of a 'would' with a 'could' *or* formulate it in terms of hyper-stability and confuse an independent problem with hyper-stability about substantive moral questions as such with a problem for expressivist theories of such judgments.

## **Conclusion**



Quasi-realists need to be able to make sense of thoughts of the form, “I believe that p, but I might be mistaken” where ‘p’ is moral, even when that belief is understood by the agent to be stable. Egan’s argument that it is impossible for a quasi-realist to make sense of such judgments goes wrong by failing carefully to track the distinction between a change which is in fact an improvement and a change which the agent takes to be an improvement (Blackburn’s point). However, even granting Blackburn’s reply, an important version of the challenge remains. Even if we are careful to track this distinction, it may turn out that judging that one might be mistaken about p will, given a plausible expressivist treatment of ‘might’, turn out to involve a belief which contradicts the relevant stability belief. Fortunately, this worry can be shown to be mistaken, once we attend to the way in which epistemic modals are context-sensitive. Once a plausible expressivism-friendly account of epistemic modals is on the table, the residual challenge is revealed to rest on a failure to distinguish a ‘would’ from a ‘could’. Nor can the argument be plausibly reinstated by shifting from stability to what I have called “hyper-stability.” For in that case, there will indeed be an epistemologically vicious form of smugness in the vicinity, but it will arise entirely from the relevant form of hyper-stability itself, and not from expressivism. Either way, expressivism evades the objection from fundamental moral error.<sup>28</sup>

---

<sup>28</sup> In this article I have focused on simply making sense of thoughts of the form, “I might be mistaken about p” for some normative ‘p’. Another way of elaborating Egan’s challenge, and one which Egan himself sometimes deploys, focuses instead on propositional attitudes other than belief which one might take towards the possibility of one’s own error in the case of stable normative beliefs. For example, someone might sincerely remark, “I think I am not morally required to give all of my money to UNICEF – I *hope* I am right about that.” In general, Ecumenical Expressivism provides an elegant way of thinking about propositional attitudes other than belief. Just as one can have normative perspective/belief pairs, one can have normative perspective/desire pairs, normative perspective/intention pairs, and normative perspective/hope pairs, where these are related via their content in the same way as in the case of normative judgment. However, a full discussion of how to extend the account developed here to a full account of the various relevant propositional attitudes must await another day, as this would require a more fully worked out account of such propositional attitudes more generally and this is a very difficult topic in its own right. It is, though, in my view, no accident that it is most natural to focus on hope in this context because hoping that p plausibly entails some level of uncertainty as to whether p is the case. On at least one view, of hope, it might be reducible to a belief that the object of the hope is epistemically possible but not epistemically necessary

---

plus a desire that the object of the hope is in fact realized. If that is right, then it should be easy to see how the account of the relevant beliefs about epistemic possibility developed here could be fed into such a reductive account. One might, for example, understand such hopes as desires that none of the epistemically possible scenarios in which there is a series of improving changes which would lead you to revise your belief that *p* is actual. Equally clearly, though, it would take me too far afield to discuss whether hope is amenable to such a reductive treatment – not to mention the subtleties of dealing with the other relevant propositional attitudes.